International Journal on Recent and Innovation Trends in Computing and Communication
Volume: 5 Issue: 1
ISSN: 2321-8169
252 – 260

Clustering of Nodes in Layered-Tree Topology for Wireless Sensor Networks

Kyung Tae Kim
College of Information and Communications Engineering
Sung Kyun Kwan University
Suwon, Korea
kyungtaekim76@gmail.com

Hee Yong Youn
College of Information and Communications Engineering
Sung Kyun Kwan University
Suwon, Korea
youn7147@skku.edu

Abstract— Wireless sensor network is composed of a large number of sensor nodes of limited energy resource. The node clustering approach can improve the scalability and lifetime of wireless sensor network. In this paper we propose a novel node clustering protocol based on layered-tree topology for self-organizing distributed wireless sensor networks. It decides optimal number of clusters by employing a new approach for setting threshold value, including the probability of optimum number of cluster-heads and residual energy of the nodes. We also introduce a new scheme for layered-tree construction in each cluster. As a result, the proposed scheme can significantly improve the energy efficiency of the network and increase its lifetime. Computer simulation shows that the proposed scheme effectively reduces and balances the energy consumption of the nodes, and thus significantly extends the network lifetime compared to the existing schemes.

***Keywords**-Cluster-head selection; Energy efficiency; Layered-tree; Network lifetime; Wireless sensor networks.*

*****

I. INTRODUCTION

Due to the development of Micro-Electro-Mechanical Systems (MEMS) tiny sensor nodes having data processing and communication capability on top of sensing circuit are available. When distributed in the target area, they form a network that can sense the environment and even react to it. The wireless sensor network (WSN) is suitable for a wide range of civil and military applications - target field imaging, intrusion detection, weather monitoring, security, and tactical surveillance and disaster management [1-2]. WSN is composed of a large number of sensor nodes and a base station (BS). The nodes are usually deployed randomly in the region of interest. The BS is engaged to give commands to the sensor nodes and gather information from them [3-4]. Typically, each sensor node consists of four basic units; sensing unit, processing unit, radio unit, and power unit. With their capabilities for monitoring and control, the network can provide a fine global picture of the target area through the integration of the data collected from many sensors each providing a limited local view [5-6].

Since sensor nodes have limited power supply which cannot be easily recharged or replaced, their operation needs to be energy efficient. In WSN the sensed data are collected in the BS usually via some intermediate node. If some sensor nodes consume all the energy, thus, the WSN cannot guarantee reliable operation due to partition of the network. The limited energy in each node affects the lifetime of the entire network, and as a result energy efficiency has been a critical design problem for the protocols and algorithms developed for WSN. Various protocols have thus been developed to efficiently organize and operate the sensor network with energy as the primary criterion [7-9].

Among various approaches proposed for energy efficient WSNs, cluster-based routing protocols have shown to be more scalable and energy-aware for WSN [10-11]. In cluster-based routing, the sensor nodes in the network are grouped according to specific requirements or metrics. Each cluster has a leader referred to as cluster-head (CH) and the member nodes. The CHs further forms hierarchical links to the BS. The clustering approach allows scalability, energy efficiency and thus longer lifetime of the network. This is due mainly to the fact that most of the sensing, data processing and communication activities can be performed within clusters. However, energy consumption of CH is significantly larger than that of other sensor nodes because CH is responsible for delivering the data aggregated by the member nodes in the cluster to the BS. Therefore the role of CH needs to be rotated among all the nodes.

In this paper we propose a novel node clustering protocol based on layered-tree topology for self-organizing distributed wireless sensor networks. The proposed scheme introduces a novel probability function, which is closely related to energy level of nodes, round information, selected count as a CH, to optimize the selection of CH. For this, we analyze the cluster number variability and find that it exist a slight inaccuracy in the computation of the node self-selected probability per round in LEACH. Also, the proposed protocol introduces a new approach to maximize the network lifetime by tree construction in a cluster. After the clusters are formed, a tree of nodes is constructed within each cluster with the CH as the root of it. All member nodes in a cluster are arranged in *m* levels starting from a CH. The member nodes set path for data transmission using level value and signal strength. Each node in a cluster sends the sensed data to their parent node, whereas aggregates them to reduce the amount of data transferred. The CH fuses the data received from the member nodes, and then transmits them to the BS. The CH selection and tree construction occur in each round of operation. As a result, the proposed scheme can significantly reduce energy consumption and increase the lifetime of the network compared to the existing schemes. Also, through the evenly distributed of CHs, the proposed scheme balances the energy consumption well among all sensor nodes. The simulation results demonstrate that proposed scheme extend more effective in prolonging the network lifetime compared with existing schemes.

The remainder of the paper is organized as follows. Section II presents the related work and Section III introduces the proposed scheme and analyzes the energy efficiency. Section IV evaluates the performance by simulation. Finally, Section V concludes the paper and outlines future research direction.

II. Related Works

A. Cluster-based Routing Protocol

As the need for efficient use of WSNs covering a large region grows dramatically, various clustering protocols have been developed to meet the operational requirements including increased network lifetime, communication latency, scalability, etc. [12-13]. The most significant problem here is how to build the clusters which reduce the energy consumption and prolong the network lifetime.

One of the most popular clustering protocols proposed for WSNs is LEACH (Low Energy Adaptive Clustering Hierarchy) [14-15]. It is a dynamic clustering protocol based on randomly deployed homogeneous stationary sensor nodes, and it serves as the basis for other clustering protocols proposed for WSNs. In LEACH, a few nodes are randomly selected as CHs per round, and the task of being a CH is rotated to balance the energy consumption. The operation of LEACH is divided into rounds, where a round begins with set-up phase followed by steady-state phase. In the set-up phase, CHs are selected and clusters are organized. In the steady-state phase, the actual data transmissions to the BS take place. After the steady-state phase, the next round begins. The CHs are stochastically selected such that each node determines a random number between 0 and 1, and if the number is less than a threshold, the node becomes a CH in that round. The threshold of node-*n*, *T*(*n*), is set as follows:

$$T(n) = \begin{cases} \dfrac{k}{N - k \times (r \bmod \frac{N}{k})} & , \ \forall n \in G \\ 0 & \forall n \notin G \end{cases} \quad (1)$$

Here *N* is the total number of nodes, *k* is the expected numbers of clusters, *r* is the number of the current round and *G* as the set of nodes that have not been CHs in the last *N*/*k* rounds. Using this approach, each node will be a CH once within *N*/*k* rounds. After *N*/*k* rounds, all nodes are once again eligible to become CHs. In LEACH, the optimal number of CHs is estimated to be about 5% of the total number of nodes. After some CHs are selected, other nodes join a cluster whose CH transmits a signal of the highest power. Each CH then sets up a TDMA schedule for all member nodes in its cluster. During the steady state phase, the member nodes send the data they generate to the CH according to the TDMA schedule. The CH aggregates and compresses the data before passing it to the BS. At the end of a round, a new set of nodes become the CHs for the subsequent round and the whole procedure repeats.

B. Probabilistic Problems of Cluster-head Selection

There exist two reasons why the energy dissipation in the cluster-based schemes employing the probabilistic CH selection approach could be unbalanced. The first is that the number of clusters is not a constant and it consumes different energy. The second is that clusters always do not ensure a proper distribution during every round. In this section we discuss these problems.

The conventional approaches for routing in WSN such as Minimum Transmission Energy and direct-transmission do not allow even energy dissipation throughout a network. To resolve this problem, LEACH employs the approach of random rotation of CHs for even energy consumption among all the nodes in the network. The LEACH requires no information on the position of the nodes and lets every node be nominated as a CH only once in *N*/*k* rounds. We define the such *N*/*k*(=1/*p*) rounds as a *cycle*.

During the setup phase of LEACH, the nodes elect themselves as CHs with pre-determined probability, *p*. LEACH fixes the number of clusters according to some factors such as network topology and communications.

However, due to the stochastic operation, different number of CHs is selected in each round. Fig. 1 shows the simulation results on a 100 node network with the k value of 5. Different number of CHs will cause the number of nodes in each cluster different, and the nodes dissipate energy unevenly in each round. We next analyze the variability in the number of clusters, and find that there exists a problem in the approach of self-elected CH per round using a problem presented in [16].

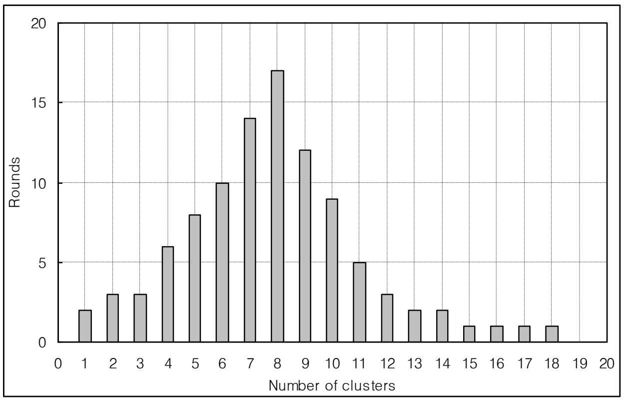


Figure 1. The distribution of the number of cluster-heads in LEACH

We assume that there are *N* nodes randomly and uniformly distributed in the target area, and the optimal value of the number of clusters is *k*. Thus each node elects itself as CH at the beginning with probability *p*(=*k*/*N*). Without losing generality, we assume (*N*/*k*) is an integer, *R*, and a *cycle* consists of *R* consecutive rounds. For simplicity, we only discuss the operation of one *cycle*.

Define *N*(*r*) (*r*=1, 2, …, *R*) as the number of candidates of CH before the cluster selection begins at round *r*, and *H*(*r*) the number of CHs after the cluster selection at round *r*. Then {*H*(*r*), *r*=1, 2, …, *R*} is a stochastic series representing the number of CHs in a *cycle*.

According to LEACH, the node should choose to become a CH at round r with probability, *p*(*r*),

$$p(r) = \left(\frac{k}{N - k \times r}\right) = \left(\frac{p}{1 - p \times r}\right) \quad (2)$$

Obviously, the distribution of the number of CHs at round *r* follows a binomial distribution, *B*(*N*(*r*), *p*(*r*)):

$$p(H(r) = k) = {}_{N(r)}C_k \left(p(r)\right)^k \left(1 - p(r)\right)^{N(r)-k} \quad (3)$$

Therefore, the mean of *H*(*r*), *η*(*r*), is

$$\begin{aligned} \eta(r) = E(H(r)) &= \sum_{i=0}^{N(r)} i \times p\{H(r) = i\} \\ &= \sum_{i=0}^{N(r)} i \times C_{N(r)}^{i} \left(p(r)^{i}\right)\left(1 - p(r)\right)^{N(r)-i} \\ &= N(r) \times p(r) \end{aligned} \quad (4)$$

The CH selection algorithm is a distributed algorithm, and each node does not know the value of $N(r)$ at each round. The algorithm adopts (N-$k \times r$) as an approximate value of $N(r)$. Thus

$$\eta(r) = N(r) \times p(r) = (N - k \times r) \times \left(\frac{k}{N - k \times r} \right) = k \quad (5)$$

Equation (5) implies that the expected number of clusters per round is k. According to [16], the optimal value of k needs to be pre-determined in order to ensure that the total energy of the network is minimized. Note, however, that $N(r)$ in equation (5) is an approximate value. With H_0=0, the actual value of $N(r)$ is

$$N(r) = N - \sum_{i=0}^{r-1} H_j \neq N - k \times r \quad (6)$$

The probability with which each node elects itself as CH is $p(r)$. Thus, the mean of $H(r)$ is

$$\eta(r) = N(r) \times p(r) = (N - \sum_{j=0}^{r-1} H_j) \times \left(\frac{k}{N - k \times r} \right) \neq k \quad (7)$$

Therefore, the expected number of clusters per round is not a constant in fact. Also, the LEACH protocol forms the clusters randomly, leading to uneven distribution of the size of the clusters. In LEACH, all CHs can be located near the edges of the network or adjacent nodes can become CHs. Some CHs may locate far from the member nodes. These phenomena can cause increased energy consumption and degrade the overall performance of the network.

As mentioned above, the mechanism of probabilistic CH selection may not ensure a proper distribution of clusters. In this paper we enhance the energy efficiency of the network by properly resolving this problem.

III. THE PROPOSED SCHEME

We first present the system model and energy model employed in the proposed approach.

A. System and Energy Model

We consider a WSN of a number of sensor nodes distributed randomly in the target area. The sensing nodes periodically form the clusters and have enough transmission power to reach the BS. The following assumptions on the sensor nodes and underlying network are usually made:

- All the sensor nodes are randomly distributed in a 2-dimensional plane according to a homogeneous spatial Poisson process with λ intensity.
- All nodes are homogeneous and have the same capabilities.
- All sensor nodes are started with the same initial energy.
- A routing and MAC infrastructure are in place, and the communication environment is contention and error-free.

We employ the radio model used in the previous schemes [14-15], [17], which is the first order radio model. To estimate the amount of energy consumption of a node, both the free space (d^2 power loss) and multipath fading (d^4 power loss) channel model are used which are based on the distance between the transmitter and receiver. For relatively short distances, the propagation loss is assumed to be inversely proportional to d^2, whereas for longer distances, it is d^4. Thus, the transmission power must be controlled to compensate this loss and thus ensure acceptable signal power at the receiver. To transmit an l-bit packet for a distance d, the radio expends the following energy.

$$E_{Tx}(l,d) = E_{Tx\text{-}elec}(l) + E_{Tx\text{-}amp}(l,d) = \begin{cases} lE_{elec} + l\varepsilon_{fs} d^2, & d < d_0 \\ lE_{elec} + l\varepsilon_{mp} d^4, & d \geq d_0 \end{cases} \quad (8)$$

When a node receives l-bit data, the energy it consumes is

$$E_{Rx}(l) = E_{Rx\text{-}elec}(l) = lE_{elec} \quad (9)$$

Here E_{elec} is the energy cost for transmitting or receiving one bit data and ε_{fs} is the amplifier coefficient. d^2 energy loss model is used for the channel attenuation. For simplicity of calculation, the transmission range of each node is assumed to be same on one condition that the transmission range should cover all the neighbors in the network. For the experiments presented in this paper, all data packets are assumed to contain same number of bits.

B. The Proposed Approach

In this subsection we introduce the proposed approach employing the tree-based clustering approach in a cluster to evenly distribute the energy load among the nodes in the network. The proposed scheme consists of two phases, clustering phase and data transmission phase. We first introduce a model deciding the optimal number of clusters in the proposed scheme.

1) Optimal number of Clusters

The cluster-based protocols adopting a simple hierarchical path selection approach do not need any information on the location of the nodes or upper layer control is not required. Since each node can be selected as CH with the same probability, the network load is relatively balanced. However, some problems need to be resolved as follows. Firstly, the optimum number of clusters, k_{opt}, needs to be decided. If the number of clusters is smaller than k_{opt}, some nodes may exhaust its energy for transmitting data to the CH locating far. With excessive number of clusters, on the other hand, the nodes will quickly deplete their energy for direct communication to the BS. LEACH sets k_{opt} as 5% of the nodes without any formal model [15]. Secondly, each node has equal probability to be a CH. If a node of low energy is selected as a CH, however, it will quickly deplete its energy due to the heavy load of CH. This shortens the network lifetime. Therefore, we need to introduce a new threshold value, which is decided based on the probability of optimal number of CHs and the residual energy of the nodes, to properly select the CHs.

During one round of operation, the energy consumption of a CH, E_{CH}, is due to three factors: the energy for data reception, data aggregation, and transmission of the aggregated data to the BS. Since the distance between the CH and BS is usually long, the multi-path (d^4 power loss) model is used, and ECH becomes:

$$E_{CH} = lE_{elec}N_1 + lE_{DA}(N_1 + 1) + lE_{elec} + l_{mp}d_{toBS}^4 \quad (10)$$

Here l is the number of bits in each data packet. N_1 is the number of member nodes in a cluster having Poisson distribution, E_{DA} is the energy cost of data aggregation, d_{toBS} is the distance from the CH to the BS. With full data aggregation, each CH needs to process n/k_{opt} signals of length-l. The average number of member nodes in each cluster is [18]:

$$E\left[N_1 \mid N=n\right] \approx E\left[N_1\right] = \lambda_0 / \lambda_1 \tag{11}$$

Here λ_0 and λ_1 denote the density of CHs and member nodes respectively. $\lambda_1 = p\lambda$, $p=k/n$, and $\lambda(=\lambda_0+\lambda_1)$ is the density of Poisson distribution process. $n=(\lambda \times A)$ is the number of nodes, whereas A is the size of the target area of square shape where the sensor nodes are deployed, k is the number of clusters, p is the probability for a node to be CH.

Without loss of generality, assume that the BS locates at the center of the target area. Hence the average distance from each CH to the BS for a square area of unit side length is:

$$\begin{aligned} E\left[D_1 \mid N=n\right] &= \iint D_1 \cdot P_A d_x d_y \\ &= \int_{-a}^{a}\int_{-a}^{a}\sqrt{x^2+y^2}\,\frac{1}{4a^2}d_x d_y = 0.765 \end{aligned} \tag{12}$$

Here D_1 the variable of Poisson distribution denoting the distance from the CH whose coordinate is (x,y) to the BS. P_A is the probability density of CH in the area A.

According to equation (11) and (12), equation (10) can be expressed by:

$$E_{CH} = \frac{n-k}{k} lE_{elec} + \frac{n}{k} lE_{DA} + lE_{elec} + 0.342a^4 l\varepsilon_{mp} \tag{13}$$

Because each member node needs to transmit only l-bit data to its CH and the distance between them, d_{toCH}, is relatively short, the free space model is used. Thus the energy used in each member node is:

$$E_{non\text{-}CH} = lE_{elec} + l\varepsilon_{fs} d^2_{toCH} \tag{14}$$

d_{toCH} is [18]:

$$E\left[L_1 \mid N=n\right] \approx E\left[L_1\right] = \lambda_0 / 2\lambda_1^{3/2} \tag{15}$$

Here L_1 is the variable of Poisson distribution signifying the sum of the distance from the member node to the CH. Then, the average distance from the member node to its CH, $E[H_1]$, is:

$$E\left[H_1 \mid N=n\right] = \frac{E\left[L_1 \mid N=n\right]}{E\left[H_1 \mid N=n\right]} = 0.5\left(\frac{k\lambda}{n}\right)^{-\frac{1}{2}} \tag{16}$$

Therefore,

$$E_{non\text{-}CH} = lE_{elec} + \frac{n}{4k\lambda} l\varepsilon_{fs} \tag{17}$$

The energy dissipated in a cluster during a single round, $E_{cluster}$, and the total energy consumption, E_{total}, are:

$$E_{cluster} = lE_{CH} + N_1 E_{non\text{-}CH} \tag{18}$$

$$E_{total} = kE_{cluster} \tag{19}$$

According to equation (13), (17), and (18), equation (19) can be expressed by:

$$E_{total} = l\cdot\left[\left(2n-k\right)\right]E_{elec} + nE_{DA} + 0.342d^4 k\varepsilon_{mp} + \frac{n(n-k)}{4k\lambda}\varepsilon_{fs} \tag{20}$$

Setting the derivative of E_{total} with respect to k to zero,

$$\frac{d}{dk}E_{total} = -E_{elec} + 0.342a^4\varepsilon_{mp} - \frac{n^2}{4k^2\lambda}\varepsilon_{fs} = 0 \tag{21}$$

Because $n = \lambda = \lambda(4\lambda^2)$, $\lambda = \dfrac{n}{4a^2}$. The optimum number of cluster, k_{opt}, and optimal probability for any node to be CH, p_{opt}, are then

$$k_{opt} = \sqrt{\frac{na^2\varepsilon_{fs}}{0.342a^4\varepsilon_{mp} - E_{elec}}} \tag{22}$$

$$p_{opt} = \frac{k_{opt}}{n} = \sqrt{\frac{a^2\varepsilon_{fs}}{n\left(0.342a^4\varepsilon_{mp} - E_{elec}\right)}} \tag{23}$$

2) *Selection of Cluster-head*

Random selection of CH causes the remaining energy of the nodes unbalanced and as a result reduces the lifetime of the network. In order to solve this problem, the CH is selected by considering the remaining energy of the nodes. This is achieved by employing a threshold, with which each node decides whether it becomes CH for the current round or not. In [17] the threshold of node-n, $T(n)$, is set as follows using k_{opt} and p_{opt}:

$$T(n) = \begin{cases} \dfrac{p_{opt}}{1-p_{opt} * \left(r \bmod \dfrac{1}{p_{opt}}\right)} \dfrac{E_{residual}}{E_{init}} k_{opt} & n \in G \\ 0 & \text{other} \end{cases} \tag{24}$$

Here G is the set of nodes that have not been CH in the last $1/p_{opt}$ rounds, $E_{residual}$ is the residual energy, and E_{init} is the initial energy. If the threshold is reduced, the probability of the nodes of high energy to be CH will increase.

Although this approach improves the network lifetime, the CH may not be able to be selected any more whereas there still exist some nodes having enough energy to transmit data to the BS. The reason for this is that the threshold of equation (24) becomes very low if the remaining energy decreases. Ultimately, the low threshold disallows the node to be selected as CH.

To deal with this problem, the proposed scheme introduces a new approach to properly select the CH. The threshold of node-n in the proposed scheme is expressed as follows:

$$T(n) = \left(\frac{p_{opt}}{1-p_{opt}\times(r \bmod \frac{1}{p_{opt}})}\left(\frac{E_{residual}}{E_{init}}\right)\right)\left(\frac{p_{opt}\sqrt{r}}{C_{ch} \bmod \frac{1}{p_{opt}}}+1\right) \tag{25}$$

In equation (25), the right-hand term is newly employed compared to equation (24). This approach ensures that every node becomes a CH exactly once within $1/p_{opt}$ rounds. After CHs are selected, each CH broadcasts an advertisement message (*ADV_Msg*) to other nodes. When a node receives an *ADV_Msg* messages from the CHs, it sends the join-request message (*Join_REQ*) to the CH which it chooses as its CH based on the received signal power. After the CH receives the message (*Join_REQ*), the CHs identify their member nodes using the received *Join_REQ* messages. Once the clusters are created, the proposed layered-tree is configured in each cluster.

3) Layered-tree configuration

After the clusters are formed, a tree is constructed in each cluster with the member nodes where the CH becomes the root. The tree construction consists of two steps. The first step is the determination of the tree level of each member node in the cluster, and the second one is the formation of the tree based on that. The proposed layered-tree structure is shown in Fig. 2.

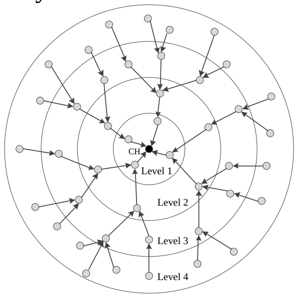


Figure 2. The tree structure in a cluster of the proposed scheme.

In tree structure of the proposed scheme, one sensor node has one parent and several children nodes. Here the CH is the root of the tree at level 0; the nodes one hop away from the CH are at level 1; the nodes two hop away are at level 2 and so on. Each node selects a parent node of the one-level inner layer which receives and fuses the data transmitted from its children nodes. Note that the number of layers in a cluster is preset based on the node density in the monitored area considering the communication distance among the nodes. Consequently, the distance between the tree layers, called communication radius (*CR*), also decided a priori.

In [19], the critical transmission range for multi-hop connectivity was presented. The authors assume that the nodes are uniformly distributed in the field and that each cell of size $C \times C$ in the network contains at least one node. In this case, the network is guaranteed to be connected if the transmission range, R_t, is $(1+\sqrt{5})C$. A cell in this context is defined as an area in the 2-dimensional space in which every node can communicate with every other node residing in every neighboring cell. In a clustered network, a cell can be defined as an area where every node can reach every other node residing in the same cell. According to [19], we obtain that *CR* is equal or longer than $(1+\sqrt{5})(\sqrt{L \times L/N})$ where L is the length of a side of a square filed.

Each member node computes its tree layer as follows.

$$\text{Tree level value} = \left\lceil \frac{D_{CH}}{CR} \right\rceil + 1 \tag{26}$$

Here D_{CH} is the distance to its CH based on the power of a signal received from the CH. D_{CH} can be obtained using the signal strength indicator (RSSI) model, which is a standard feature in wireless radios and attracted a lot of attentions in recent literature. It is defined as the voltage of the signal measured by the receiver circuit. RSSI can be used to estimate the distance from a transmitter to a receiver. The principle of RSSI ranging techniques describes the relationship between the received power, P_r, transmitted power of the signal, P_t, and the distance among wireless sensor nodes, d, which is shown in equation (27).

$$P_r = P_t \times (1/d)^n \tag{27}$$

Here n is the transmission factor whose value depends on the environment of the propagation [20]. The received power in *dBm* is then given as

$$P_r(dBm) = A - 10n\log d \tag{28}$$

In equation (28), notice that the value of parameter A and n determine the relationship between the strength of received signal and the distance of signal transmission. The distance between the source and receiver can be found out by the strength of the radio signal received.

$$d = (xP_t / P_r)^{1/2} \tag{29}$$

Here, d is the distance between the transmitter and the receiver.

We can also calculate the maximum tree value (L_{max}) using the longest link of the field, L_l, and CR as follows.

$$L_{max} = \frac{L_l}{CR} + 1 \tag{30}$$

After the leveling of the member nodes is completed, a tree is constructed. For the tree construction, L_{max}, number of iterations (steps) is taken. Every step should be long enough for the nodes residing in that tree layer to receive and transmit messages between the parent and children nodes. The path setup starts from the root and terminates at the nodes at layer-L_{max}.

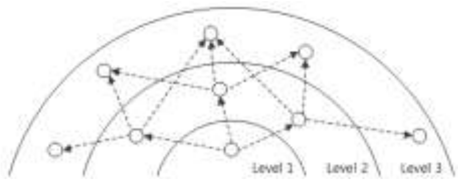


Figure 3. Broadcast for tree construction in a cluster.

During the tree construction, the nodes broadcast, using CSMA, a message which is composed of the node's ID, layer number, and ID of the parent node. Fig. 3 shows the broadcast process for tree construction in a cluster. Here each node decides a node as its parent node that transmits a signal of the strongest power. When a node receives the reply from a node one level outer layer, it lists it as a child node. Then they turn off their radio until the data collection begins. The nodes of which tree level is 1 elect the CH as their parent node.

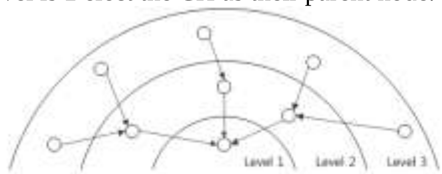


Figure 4. The final paths of tree layout.

The tree construction process continues until the nodes at outermost layer decide their parent node. Fig. 4 shows the final path of data transmission between the nodes after the tree construction is over for the nodes of Fig. 3.

4) Data Collection and Transmission

Once layer-based tree construction is complete in each cluster, the data collection and transmission phase starts. In this phase, each node sends the collected data to the parent node during the pre-allocated time slot appointed by the CH. We choose to implement a time division multiple access (TDMA)-based MAC layer for the slot assignment. The TDMA MAC layer provides two advantageous features as clock synchronization is built in the TDMA protocol and collision among the nodes in the cluster can be avoided by assigning non-overlapping time slots.

When the nodes receive the TDMA schedule and the assigned frequency band, they perform data fusion and transmission in the time-slot allotted to them. In every slot one or more node transmits gathered data to the lower level nodes, from the outermost layer toward the innermost layer nodes. When the data from the nodes of all levels in the cluster have been received, the CH sends the fused data to the BS. Assume there exist n nodes in a cluster, excluding the CH. In the existing cluster-based schemes where all member nodes send the data to the CH, n time-slots are required and the time complexity is $O(n)$. In the proposed scheme, however, the time depends on the number of levels of the tree. This is because all nodes in each layer operate simultaneously.

C. Analysis of Energy Consumption

As outlined in Section B, the total energy consumed per round is due to two operations of node clustering and data transmission. The following subsections provide the mathematical models estimating the energy consumption of each operation.

1) Node Clustering

As described in Section B, node clustering in tree configuration occurs in each round. For the cluster formation, first, each CH transmits *ADV_Msg* to other nodes in the network. The nodes may receive the message from one or more CHs and select a one based on the received signal power. Finally, each node sends a *Join_REQ* message to the CH. The energy consumed by CH during this phase is given by equation (31).

$$E_{Ch\_Sel\_clustering} = lE_{elec}N_1 + \left(lE_{elec} + l\varepsilon_{mp}d^4\right) \quad (31)$$

And the energy consumed by each member node is given by equation (32).

$$E_{Non\text{-}Ch\_Join\_clustering} = \left(lE_{elec} + l\varepsilon_{fs}d^2\right) + E_{Rx}(l) \quad (32)$$

Therefore, the energy consumed during CH selection and clustering is given by

$$E_{C\text{-}Phase} = E_{Ch\_Sel\_clustering} + N_1 E_{Non\text{-}Ch\_Join\_clustering} \quad (33)$$

In the proposed scheme the nodes in each cluster are connected in tree topology where the CH is the root. In the tree construction phase, the nodes in each level handle one transmission process and one reception process. The transmission process is for broadcasting *path_message* for path setup and the reception process is for receiving *path_message* from neighbor nodes, respectively. Therefore, the energy consumed in the member node during tree construction, $E_{Non\text{-}Ch\_tree}$, is given by equation (34).

$$E_{Non\text{-}Ch\_tree} = p_m\left(E_{elec} + \varepsilon_{fs}d^2\right) + E_{Rx}(p_m) \quad (34)$$

Here p_m is the length of *path_message*. Once tree construction is complete, the CH creates a TDMA schedule for all the member nodes in its cluster and sends it to them. Therefore, the energy consumed by CH during clustering phase can be expressed as equation (35) using equation (31):

$$E_{Ch\_clustering} = E_{Ch\_Sel\_clustering} + E_{schedule} + \left(lE_{elec} + l\varepsilon_{fs}d^2\right) \quad (35)$$

The member nodes in a cluster receive TDMA schedule from the CH. Therefore,

$$E_{M\text{-}node\text{-}tree} = p_m\left(E_{elec} + \varepsilon_{fs}d^2\right) + E_{Rx}(p_m) + E_{Rx}(l) \quad (36)$$

The energy consumption of non-CH node during the clustering phase, $E_{Non\_Ch\_clustering}$, is given by summing up equation (32) and (36).

$$E_{Non\text{-}Ch\_clustering} = p_m\left(E_{elec} + \varepsilon_{fs}d^2\right) + E_{Rx}(p_m) + E_{Rx}(l) + \left(lE_{elec} + l\varepsilon_{fs}d^2\right) + E_{Rx}(l) \quad (37)$$

Here, $lE_{elec} + l\varepsilon_{fs}d^2$ can be expressed as equation (38) because of the short-distance transmission within a cluster.

$$lE_{elec} + l\varepsilon_{fs}d^2 = \left(lE_{elec} + l\varepsilon_{fs}\frac{M^2}{2\pi k}\right) \quad (38)$$

The amount of energy consumption in a single cluster during the clustering phase is represented as follows.

$$\begin{aligned} E_{Cluster\_clustering} &= E_{Ch\_clustering} + N_1 E_{Non\text{-}ch\_clustering} \\ &= E_{schedule} + l\left(E_{elec} + \varepsilon_{fs}\frac{M^2}{2\pi k}\right) \\ &+ N_1\left(lE_{elec}\right) + l\left(E_{elec} + \varepsilon_{mp}d^4\right) \\ &+ p_m\left(2\left(E_{elec}\right) + \varepsilon_{fs}\frac{M^2}{2\pi k} + \right) \\ &+ l\left(\left(lE_{elec} + l\varepsilon_{fs}\frac{M^2}{2\pi k}\right) + 2E_{elec}\right) \end{aligned} \quad (39)$$

The total energy consumption in a network during the clustering phase in a round can be obtained by multiplying equation (39), expressing the amount of energy consumption in a single cluster, by the number of clusters, k_{opt}.

$$E_{Clustering} = k_{opt}\left(E_{Ch\_clustering} + N_1 E_{Non\text{-}ch\_clustering}\right) \quad (40)$$

2) Data Transmission

In the data transmission phase, all non-CH nodes send a single data message of l bytes to the parent node and the nodes except the leaf nodes receive data from their children nodes. In equation (41), the term (n-k_{opt}) represents the number of non-CH nodes and n_L is the number of member nodes in each cluster. The CH sends an aggregated data to the BS.

$$E_{Non\text{-}Ch\_trans} = l\left(n - k_{opt}\right)\left(E_{elec} + \varepsilon_{fs}d^2\right) + lE_{elec}\left(n - k_{opt} - k_{opt}(n_L)\right) \quad (41)$$

In equation (42) the first term represents the amount of energy consumed in the transmission from the CH to the BS and the second term accounts for the energy consumed by the CH in receiving the data from its children nodes.

$$E_{Ch\_trans} = lk_{opt}\left(E_{elec} + \varepsilon_{mp}d^4\right) + k_{opt}\left(lE_{elec}\left(n_{Dn} + E_{DA}\right)\right) \quad (42)$$

Therefore, the total energy consumed by all nodes during the data transmission phase in a round is expressed by;

$$E_{dataphase} = E_{Ch\_trans} + E_{Non\text{-}ch\_trans} \quad (43)$$

 ISSN: 2321-8169
252 – 260

IV. Performance Evaluation

In this section we evaluate the performance of the proposed scheme using a simulator written C++, and compare it with the existing protocols. The simulation evaluates the network lifetime, number of messages received, amount of energy spent by the CHs per round, and the number of nodes alive. The simulator also evaluates residual energy of each node and the communication paths between the nodes of the network. Table I lists the parameters used in the simulation.

TABLE I. THE PARAMETERS USED IN THE SIMULATION

Parameter	Value
Network model	
Network size	$100m \times 100m$, $200m \times 200m$
Location of BS	Center of target area
Number of nodes	100
Data packet size	500 byte
Packet size for path set-up	20 byte
Number of frames per round	1
Network topology	Random
Energy model	
Initial energy of each sensor	1.0J
E_{elec}	50 nJ/bit
ε_{fs}	10 pJ/bit/ m^2
ε_{mp}	0.0013 pJ/bit/ m^4
E_{DA}	5 nJ/bit/signal
Path loss exponent (m)	2

In the simulation the BS is assumed to be fixed and located at the center of the target area, and it has unlimited resources including the power. In order to study how the proposed scheme works with the network of different densities, the simulations are conducted with two different sizes of network of $100m \times 100m$ and $200m \times 200m$. 1000 simulation runs are executed, and the results are averaged. In the simulations the radio model presented in Section A is adopted. The free space propagation model is used when the propagation distance is smaller than the threshold distance d_0. Otherwise, the ground reflection (two-ray) propagation model is used [14], [15], [17]. For the simplicity, an ideal MAC layer and error-free communication links are assumed.

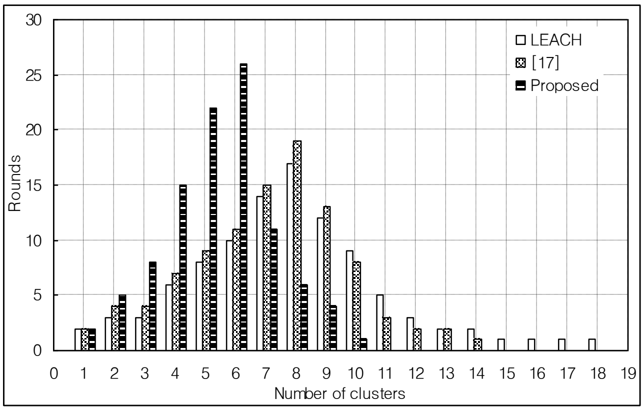


Figure 5. The distribution of the number of clusters with different schemes

To investigate the effectiveness of clustering, we first examine the distribution of the number of clusters with LEACH, [17], and the proposed scheme. In Fig. 5, it is apparent that the number of clusters with the proposed scheme is more uniform than with LEACH and equation (24). LEACH employs a fully random approach to decide the CHs, resulting in a large distribution of the number of clusters as shown in the figure. Since [17] utilizes the remaining energy of each node in CH selection, it ensures lower variance of the number of clusters than LEACH. In addition to the remaining energy in each node, however, the proposed scheme considers other factors including the count elected as CH. As a result, it allows more uniform number of clusters than LEACH and [17].

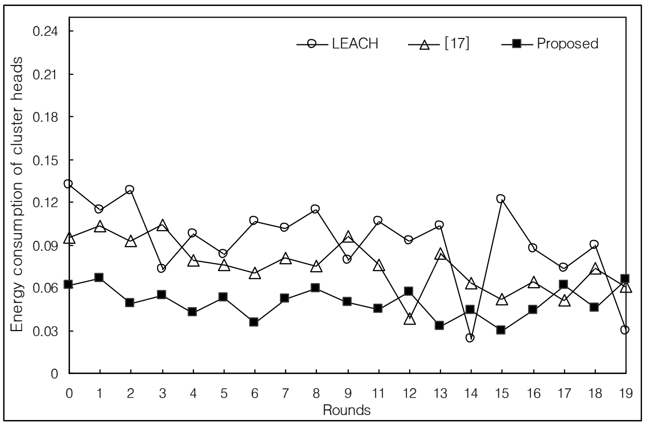


Figure 6. The amount of energy spent by the cluster-heads per round

Next, the amount of energy spent by the CHs is investigated. When all nodes are alive, the amount of energy spent by all CHs for 20 rounds of simulations is shown in Fig. 6. It shows that the energy consumed by the CHs per round in the proposed scheme is much lower than with others. If elected number of CHs is smaller than the optimal number of cluster, k_{opt}, the size of the clusters in the network will be large. Therefore, the CHs have to consume large energy. If the number is higher than k_{opt}, on the contrary, the energy efficiency is reduced because many CHs need to transmit the aggregated data to the BS. In the proposed scheme, the number of CHs is decided properly, and thus shows higher energy efficiency than others.

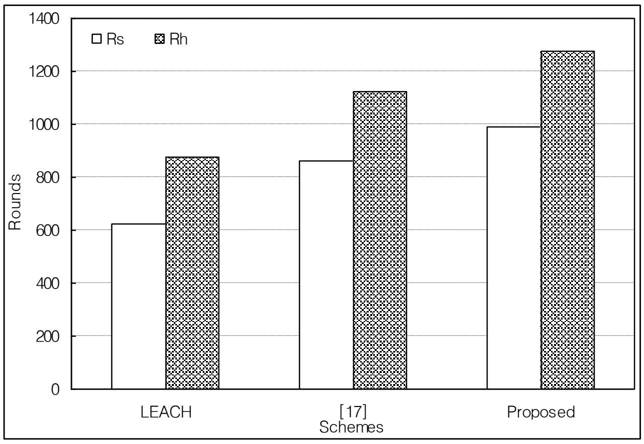
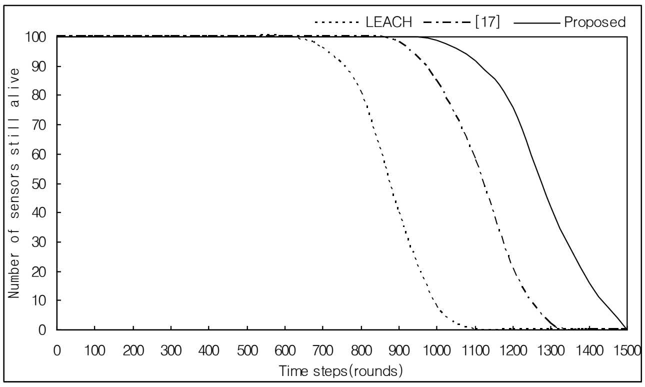


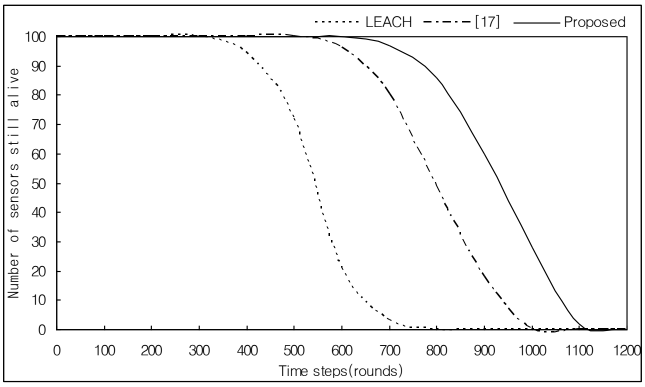
Figure 7. The comparison of R_s and R_h of the schemes

To evaluate the network lifetime of the schemes, two metrics, the round a node starts to die (R_s) and the round half of the nodes are dead (R_h) are shown in Fig. 7. Notice that R_s of the proposed scheme is approximately 74% and 19% higher than that of LEACH and equation (24), respectively. R_h is approximately 61% and 18% higher.

Fig. 8 shows the number of sensor nodes still alive over the simulation time. The proposed scheme clearly improves the total network lifetime over others. In Fig. 8 (b), 100 nodes are randomly placed in a 200*m* × 200*m* area with the BS located at (100, 100). Fig. 8 (b) is the simulation result with this condition. Observe that the proposed protocol outperforms other protocols even for larger size network.



(a) 100*m*×100*m*



(b) 200*m*×200*m*

Figure 8. The comparison of network lifetime

Fig. 9 shows the total number of messages received by the BS over time with different schemes. For this experiment, we consider a network of 100 sensor nodes randomly distributed in a place of 100*m* × 100*m* network, where each node begins with an initial energy of 1.0J. The figure clearly illustrates the effectiveness of the proposed scheme in delivering significantly more messages than its counterparts. This was achieved by deciding optimal number of CHs and forming a layered-tree in each cluster.

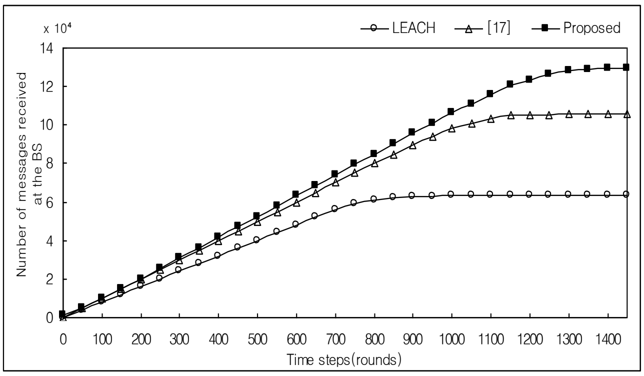


Figure 9. The Total number of message received by the BS

V. CONCLUSION

In this paper we have proposed a novel node clustering protocol based on layered-tree topology for self-organizing distributed wireless sensor networks. It decides optimal number of clusters using a new approach of threshold set-up, which employs the probability of optimum number of CHs in the network. Through the analysis of the variation in the number of clusters, the optimal number of CHs was found. Also, we have introduced a new approach of layered-tree construction in each cluster to maximize the network lifetime. As a result, the proposed scheme can significantly reduce energy consumption and increase the lifetime of the network compared to the existing schemes. Through the evenly distributed CHs, the proposed scheme balances the energy consumption well among all sensor nodes.

In the future we will continue to study to decide optimum number of layers of the tree in each cluster. In addition, we will also investigate how to maximize the network lifetime of heterogeneous network where the sensed data are not correlated and data aggregation is not possible.

ACKNOWLEDGEMENTS

This research was supported by Institute for Information & communications Technology Promotion (IITP) grant funded by the Korea government (MSIP) (No.B0717-17-0070), Basic Science Research Program through the National Research Foundation of Korea (NRF) funded by the Ministry of Education, Science and Technology (2016R1A6A3A11931385), the second Brain Korea 21 PLUS project.

REFERENCES

[1] I. F. Akyildiz et al., "Wireless sensor networks: a survey," Computer Networks, vol. 38, no. 4, pp. 393–422, 2002.

[2] M. Tubais hat and S. Madria, "Sensor networks: an overview," IEEE Potentials, vol. 22, pp. 20–23, 2003.

[3] F. Bajaber and I. Awan, "Adaptive decentralized re-clustering protocol for wireless sensor networks," Journal of Computer and System Sciences, vol. 77(2), pp. 282–292, March 2011.

[4] A. Alajlan, K. Elleithy, and V. Pande, "A New Programming Model to Simulate Wireless Sensor Networks: Finding the Best Routing Path," Journal of computer Networks & Communications, vol. 5, no. 4, pp. 1–14, 2013.

[5] R. A. Gupta and M.-Y. Chow, "Networked control system: Overview and research trends," IEEE Trans. on Industrial Electronics, vol. 57(7), pp. 2527–2535, 2010.

[6] J. A. Khna, H. K. Qureshi, and A. lqbal, "Energy management in Wireless Sensor Networks: A survey," Computer & Electrical Engineering, vol. 41, 159–176, 2015.

[7] J. Yick, B. Mukherjee, and D. Ghosal, "Wireless Sensor Network surveys," Computer Networks, vol. 52, pp. 2292–2330, 2008.

[8] N. Rathi, J. Saraswa and P. P Bhattacharya, "A Review on Routing Protocols for Application in Wireless Sensor Network," International Journal of Distributed and Parallel System, vol. 3, no. 5, pp. 39–58, 2012.

[9] S. K. Singh, M. P. Singh, and D. K. Singh, "Routing Protocols in Wireless Sensor Networks – A survey," International Journal of Computer Science & Engineering Survey, vol. 1, no. 2, pp.60–64, 2010.

[10] M. Maimour, H. Zeghilet and F. Lepage, "Cluster-based Routing Protocol for Energy Efficiency in Wireless Sensor Networks," Sustainable Wireless Sensor Networks, pp. 167–188, 2010.

[11] X. Liu, "A Survey on Clustering Routing Protocols in Wireless Sensor Network," Sensors, vol. 12, pp. 11113–11153, 2012.

[12] C. Li, H. X. Zhang, B. B. Hao, and J. D. Li, "A survey on routing protocols for large-scale wireless sensor networks," Sensors, vol. 11, pp. 3498–3526, 2011.

[13] N. Nokhanji and Z. M. Hanapi, "A Survey on Cluster-based Routing Protocols in Wireless Sensor Networks," Journal of Applied Sciences, vol. 14, pp. 2011–2022, 2014.

[14] W. R. Heinzelman, A. Chandrakasan, and H. Balakrishnan, "Energy-Efficient Communication Protocol for Wireless Micro-sensor Networks," In Proc. of the Hawaii International Conf. on System Science, Maui, Hawaii, 2000.

[15] W. B. Heinzelman, A. Chandrakasan, and H. Balakrishanan, "An Application-Specific Protocol Architecture for Wireless Microsensor Networks," IEEE Trans. on Wireless Communications, vol. 1, pp. 660–670, 2002.

[16] H. Liu, L. Li, and S. Jin, "Cluster number variability problem in LEACH," Lecture Notes in Computer Science, vol. 4159, pp. 429–437, 2006.

[17] M. C. M. Thein, and T. Thein, "An Energy Efficient Cluster-Head Selection for Wireless Sensor Networks," Proc. of Int. Conference on Intelligent Systems, Modelling and Simulation (ISMS), pp. 287–291, 2010.

[18] S. G. Foss and S. A. Zuyev, "On a Voronoi Aggregative Process Related to a Bivariate Poisson Process," Advances in Applied Probability, vol.28, pp. 965–981, 1996,

[19] O. Younis and S. Fahmy, "HEED: A Hybrid, Energy-Efficient, Distributed clustering approach for Ad Hoc sensor networks," IEEE Trans. on Mobile Computing, vol. 3, pp. 366–379, 2004.

[20] J. Xu, W. Liu, F. Lang, Y. Zhang, and C. Wang, "Distance Measurement Model Based on RSSI in WSN," Wireless Sensor Network, pp. 606–611, 2010.